DOCUMENT RESUME

ED 149 309 CS 003 883

AUTHOR Oliver, Jo Ellen
TITLE Validation of a Criterion-Referenced Instrument to Measure Thinking Prerequisites to Comprehension.
PUB DATE Dec 77
NOTE 15p.; Paper presented at the Annual Meeting of the National Reading Conference (27th, New Orleans, Louisiana, December 1-3, 1977)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Cognitive Measurement; *Concept Formation; Grade 5; Intermediate Grades; *Reading Comprehension; *Reading Tests; Test Construction; Test Reliability; Test Validity

ABSTRACT

A primary-grade-level instrument was developed to measure thinking prerequisites to reading comprehension. Levels of responses are identified as either perceptual or conceptual; the ability to synthesize concepts from several sources and make generalizations is also measured. Analyses of partial construct validity, of cross-form reliability, and of cross-reader reliability are discussed, using data from fifth-grade students. Scoring instructions and test items for Form A of the test are included as an appendix. (AA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED149309

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

VALIDATION OF A CRITERION-REFERENCED INSTRUMENT TO MEASURE THINKING PREREQUISITES TO COMPREHENSION

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Jo Ellen Oliver
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Jo Ellen Oliver
University of Georgia

A concensus of opinion as to what the reading process entails does not exist, yet the search continues for more adequate means of assessing reading comprehension.

Huey (1908), as he pondered the mysteries of the human brain, raised the question, "What is reading?" Nearly seventy years later, a summary of research in the psychology of reading concludes in this manner:

> reading as a psychological process . . . is still poorly understood in the sense that a common understanding of the reading process is accepted and translated into instructional practice. (Harris, 1969, p. 1075)

Farr (1969) has said, "The most pressing research need in measuring comprehension is for a clear understanding of the nature of reading

CS003383

comprehension" (p. 63). Farr further states, "Another research need is the development of criterion tests for measuring reading comprehension (p. 64).

Of standardized tests, Farr says, ". . . tests which are suggested by the publishers for a particular grade level may not be valid for the advanced or retarded readers of that grade" (p. 138). A further criticism of existing tests by Farr is that they reveal "what a student can do and not how he does it" (p. 215).

Recently, a revival of interest in the conceptualization process has been evident. Gibson (1976) said, "One trend (in cognitive development) which seems to be especially important is the increasing ability to be aware of one's own cognitive processes . . ." (p. 187). In calling for research MacGinitie (1975-76) stated that additional study was needed to see

> if essentially the same tasks [cognitive operations] could be presented in ways that provide fewer or simpler operations . . . we should find out what processes current reading instruction requires . . . and if they can be taught. (p. 16)

If a definition of the conceptualization processes necessary for reading comprehension is difficult to formulate, is it any wonder that assessment of comprehension is also difficult? Assessment of conceptualization is difficult because the activities involved in deriving meaning and demonstrating understanding have not been adequately defined. Elkind (1969) claimed, "Every method of concept assessment presupposes a particular theory of concept development" (p. 31).

A study based upon the concept-development theory of Henry (1974) and upon the summary of research on children's thinking done by Russell (1956) was done by this researcher. Forty-one low achieving fifth-graders who were tutoring low-achieving first graders were taught to discriminate between letters and words, and finally between concepts in children's literature. The children were taught to abstract relevant information and to make generalizations. This process was similar to Henry's (1974) four steps of (1) joining, (2) excluding, (3) selecting, and (4) implying. No existing test seemed appropriate to assess children's growth. Henry stated that no test was available to measure his method and advocated the construction of a criterion-referenced instrument (p. 11).

As Farr has stated, a model of reading comprehension is needed. As has already been discussed, the conceptual model for study previously mentioned (Oliver, 1977) is a synthesis of Henry's (1974) theory and Russell's (1956) generalizations about the nature of concept attainment.

Carver (1973) states that nearly all standardized reading tests do measure reasoning rather than reading (p. 53), but that two types of criterion-referenced tests are needed in order to do research about the relationship of reasoning to reading: (1) edumetric tests which measure reading and <u>not</u> reasoning, and (3) edumetric tests which <u>do</u> measure reasoning (p. 55). The test created for the previously mentioned study (Oliver, 1977) would probably fall into the second category.

For those who argue that reading is reasoning, however, Carver's division would seem arbitrary. Stauffer (1969), who does not separate

reading from the thinking process, sets forth some criteria to be considered in the construction of an instrument to measure thinking processes:

1. Several tasks which require the learner to commit to memory the most important pieces of specific information.
2. Several tasks which require the learner to generalize or draw conclusions.
3. Several tasks to require the student to apply newly learned generalizations to new information.
4. Several tasks which require the student to discover relationships between newly-learned information and information previously learned (p. 403).

Otto (1973) in his discussion of criterion-referenced measures and norm-referenced measures, says of criterion-reference tests . . . "they get directly at the performance of individuals with regard to specified instructional objectives" (p. 18).

The instrument discussed in this study was created to identify children's levels of responses as either perceptual or conceptual and to measure their ability to synthesize concepts from several sources and make generalizations. An assumption was made that children move from the concrete to the abstract (Parts I and II), from specific to generic (Part III), and that the ability to synthesize concepts from several sources is an important prerequisite to comprehension.

The above test was created, written at the primer level, utilizing comparisons between letters, words, and then stories. The test was tried on a number of children and then submitted to two college

professors and a seminar of 30 doctoral students in reading for scrutiny. Their suggestions were incorporated into a revised form and its alternate form. These forms were then submitted to four college professors with backgrounds in testing and child development to establish content validity. To establish partial construct validity, the test or its alternate was administered randomly to 30 fifth-grade children. These children were then given 90 minutes of training in the concept-attainment process. After eight days had elapsed, the children were given the alternate forms of the tests they had previously taken. A paired t test was performed (Popham & Sirotnik, (1973). For the total scores, the t value was 2.41 ($p < .05$). For conceptual scores only, the t value was 3.23 ($p < .005$). Mean total points increased 2.5. Mean conceptual points increased 2.6. See Table 1.

To establish cross-form reliability, 100 fifth-graders in a different school were randomly given Form A or Form B. Twelve days later, the children were given the alternate forms of the tests they had taken. Children were not given training. A Pearson product-moment correlation test (Popham & Sirotnik, p. 85) was done. Cross-form reliability for the total test score was .80 ($p < .001$). For the conceptual parts of the tests, cross-form reliability was .74 ($p < .001$). See Table A2.

Womer (1974) has said that test-retest or parallel-form types of reliability are not as important for criterion-referenced tests (39) as for norm-referenced tests. In some cases, says Womer, a measure of internal consistency can be useful in a criterion-referenced test. However, as this test is presently constructed, split-half reliability would not yield useful information. The correlations between the two

scores of the tests, conceptual and total, were .89 ($p < .001$). The format of the test also does not easily lend itself to item analysis. Womer, however, states that the establishment of items which discriminate between high- and low-achieving individuals is not essential in a criterion-referenced test, its basic goal being "to describe behavior as accurately as possible in relation to independently pre-established standards of performance deemed important to the user (developer) of the test" (p. 38).

A limitation in the use of this test is that interpretive scoring is required. After reading a large number of children's answers, some problem areas were identified. These problem areas are discussed in the scoring guides for Forms A and B. To check the reliability of this tester's scoring, a reading professor from another college was asked to randomly select 30 of the tests which had been scored and to rescore them without referring to original scores, using the criteria set forth. Of 750 available item responses, only six were not in agreement. More than 99 percent of the item scores were agreed upon. A Pearson product-moment correlation test (Popham & Sirotnik, p. 85) found cross-reader reliability to be .9958 for the scores of entire test and .9915 for the conceptual scores.

Table 1 consists of scores of 30 students tested to establish partial construct validity. Table 2 consists of scores of 100 students tested to establish cross-form reliability and Table 3 consists of 30 scores used to establish cross-reader reliability. Table 4 consists of correlations between this test and other tests administered to low-achieving fifth-grade students (Oliver, 1977). Moderate correlations

exist between the instrument and SRA-PMA (Thurstone, 1963) Total score, the SRA-PMA Reasoning and Number subtests, and the Comprehension subtest of the California Achievement Test (Tiegs and Clark, 1970) Reading. Table 5 shows the mean scores of four classes of untrained fifth-graders designated by their principal as "normal readers." Low-achieving fifth-graders, as can be seen, scored higher, after training, than "normal readers" who had not been trained. Following the Tables are copies of the test, Forms A and B, and scoring directions for each. *

Dissatisfaction with standardized tests is increasing to the extent that professional organizations, such as the National Education Association, have called for a moratorium on their use (Rudman, 1977). Although this test was created for a specific study, a more general need for similar criterion-referenced tests to describe how children think may exist.

*Forms A and B, and the scoring directions for the California Achievement Test were removed due to copyright restrictions.

REFERENCES

Carver, R. P. Reading as reasoning: Implications for measurement. Assessment Problems in Reading. Newark, Del.: International Reading Association, 1973, 44-56.

Elkind, D. Conservation and concept formation. Studies in Cognitive Development, London, Toronto: Oxford Press, 1969, 171.

Farr, R. Reading, what can be measured? Newark, Del.: International Reading Association, 1969.

Gibson, E. J. Trends in perceptual development: Implications for the reading process. In H. Singer, & R. B. Ruddell (eds.), Theoretical Models and Processes of Reading. Newark, Del.: International Reading Association, 1976, 186.

Harris, T. L. Reading. In R. L. Ebel (ed.), Encyclopedia of Educational Research, 4th ed. New York: Macmillan, 1969, 1069-1104.

Henry, G. H. Teaching reading as concept-development: Emphasis on affective thinking. Newark, Del.: International Reading Association, 1974.

Huey, E. B. The Psychology and Pedagogy of Reading. First published in 1908 by Macmillan Company. Springfield, Mass.: Massachusetts Institute of Technology Press, 1968.

MacGinitie, W. H. Research suggestions from the literature search. Reading Research Quarterly, 1975-76, 11 (1), 7-35.

Oliver, J. The effects of training in synthesis-level conceptualization upon low-achieving first and fifth graders involved in a literature-based, cross-age tutoring program. Unpublished doctoral dissertation, University of Georgia, Athens, Georgia 1977.

Otto, W. Evaluating instruments for assessing growth in reading. Assessment Problems in Reading. Newark, Del.: International Reading Association, 1973.

Popham, J. W., & Sirotnik, K. A. Educational Statistics, Use and Interpretation. New York: Harper and Row, 1973.

Rudman, H. C. "The standardized test flop." Phi Delta Kappan, Nov. 1977, pp. 179-188.

Russell, D. Children's Thinking. Boston, Mass.: Ginn and Co., 1956.

Stauffer, R. G. Teaching Reading as a Thinking Process. New York: Harper and Row, 1968.

Thurstone, G. T, Primary Mental Abilities Tests, Level 4-6, Chicago, Illinois: Science Research Associates, Inc. 1963.

Tiegs, E. W., & Clark, W. W. California Achievement Test, Reading, Level 3. Monterrey, California: CBT/McGraw-Hill, 1970.

Table 1

Paired t Test

Student	Pretest		Posttest	
	Total	Conceptual	Total	Conceptual
01	15	06	20	14
02	20	08	17	10
03	16	04	17	08
04	21	14	22	10
05	12	06	15	10
06	07	00	12	02
07	11	02	10	05
08	11	06	11	08
09	09	04	14	08
10	21	08	15	08
11	15	08	19	10
12	14	16	04	04
13	09	02	11	06
14	19	08	24	14
15	21	08	21	08
16	17	10	18	06
17	24	12	32	22
18	08	02	22	08
19	10	06	22	14
20	26	10	18	10
21	13	06	20	08
22	07	06	14	06
23	18	04	18	04
24	12	05	28	16
25	12	02	14	08
26	13	04	13	06
27	19	06	21	12
28	15	06	14	08
29	22	04	22	04
30	17	08	21	14

Table 2

Cross Form Reliability

Student	Pretest Form A or B (1) (2)	Pretest Total	Pretest Conceptual	Posttest Form A or B (1) (2)	Posttest Total	Posttest Conceptual
01	1	11	04	2	18	08
02	1	04	02	2	05	00
03	1	24	08	2	14	04
04	1	06	06	2	03	02
05	1	30	14	2	27	14
06	1	33	20	2	33	18
07	1	35	20	2	31	18
08	1	33	24	2	34	20
09	1	06	02	2	08	02
10	1	22	08	2	18	06
11	1	15	08	2	13	02
12	1	21	08	2	27	14
13	1	18	06	2	21	06
14	1	24	10	2	29	14
15	2	12	04	1	11	02
16	2	30	14	1	33	22
17	2	17	06	1	26	18
18	2	18	12	1	30	14
19	2	28	12	1	25	10
20	2	25	14	1	27	14
21	2	19	10	1	17	06
22	2	23	10	1	27	14
23	1	33	16	2	31	18
24	1	27	14	2	28	10
25	1	23	13	2	31	18
26	1	09	04	2	08	02
27	1	18	06	2	27	10
28	1	21	12	2	23	14
29	1	19	12	2	26	18
30	1	18	08	2	19	10
31	1	19	08	2	20	10
32	1	32	16	2	28	18
33	1	12	02	2	13	08
34	1	28	16	2	32	18
35	1	16	12	2	18	12

Table 3

Cross-Reader Reliability

Conceptual Scores		Total Scores	
Reader 1	Reader 2	Reader 1	Reader 2
12	12	22	22
10	10	28	28
10	10	20	20
20	20	32	32
10	10	27	27
10	10	22	22
20	20	33	33
06	06	18	18
12	12	18	18
18	18	31	31
02	02	08	08
06	06	14	14
07	07	15	15
12	12	19	19
12	12	25	25
06	06	09	09
16	16	32	32
08	08	19	19
04	04	17	17
12	12	30	30
08	08	18	18

Table 4

Results of Pearson Product-Moment Correlations Between a Tentative Criterion-Referenced Test to Measure Thinking Processes and Other Tests Taken by Fifth-Grade Subjects

	A Tentative Criterion-Referenced Test to Measure Thinking Processes	
	Conceptual Score	Total Score
SRA Verbal	.1108	.1685
SRA Number	.4580	.4421
SRA Spatial	.2078	.2531
SRA Reasoning	.4168	.4240
SRA Perceptual Speed	.3701	.4876
SRA Total	.4982	.5673
CAT--Vocabulary	.0107	.0263
CAT--Comprehension	.4019	.4513
CAT--Total	.2973	.3739
Powell Attitude Scale	.1939	.0311
Piers-Harris Children's Self-Concept Scale	.0187	.2050
CRTTMTP Conceptual Score	1.000	.8433
CRTTMTP Total Score	.8433	1.000

Table 5

A Comparison of Scores of Untrained Normal Readers, Trained and Untrained Poor Readers on A Tentative Criterion-Referenced Test to Measure Thinking Processes

Comparison Group	n		$\bar{X}$ Score Conceptual	$\bar{X}$ Score Total
A	27		9.77	20.96
B	31		10.39	20.94
C	29		8.96	19.27
D	24		6.45	16.75
Total Grade	111		8.89	19.4
		Experimental		
Tutoring and Concept Training	22	Before	3.17	8.92
		After	10.76	19.81
Tutoring	19	Before	1.64	5.27
		After	5.2	11.68